DOCUMENT RESUME

ED 282 924 TM 870 344

AUTHOR Cancelli, Anthony A.
TITLE Methods for Arriving at Clinical Judgments in Peer Evaluation.
PUB DATE Apr 87
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Ability; *College Faculty; Data Analysis; Data Collection; Evaluation Criteria; *Evaluation Methods; Higher Education; Instructional Development; Interviews; Lesson Observation Criteria; Models; *Peer Evaluation; *Teacher Behavior; *Teacher Evaluation; Teaching Methods

ABSTRACT

The peer evaluation model of university teaching is composed of three domains of behavioral and cognitive skills which, in combination, represent the elements expected of professionally competent teachers. These domains are: (1) the conception of the professional role for which students are being prepared; (2) instructional design; and (3) the execution of the instructional strategy. The measurement of competence in each of these domains poses a particularly difficult challenge. Consistent with the underlying philosophy of the model, methods were developed for collecting data for different dimensions in each of the domains. The data base for making judgments was constructed from interviews with the faculty member and his/her students, classroom observations, and a review of the course syllabus and other course materials. Independent clinical judgments based on definitions of expertness were made by each of six reviewers. Following these independent judgments, disagreements were discussed and consensus reached. The purpose of this paper is to describe the methods employed to collect the data and the processes used to arrive at a consensus clinical judgment. Appendices contain: (1) the scoring sheet used by reviewers to evaluate a professor; and (2) the scoring agreement sheet used by the chairperson to identify the percent of agreement between the reviewers. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED282924

Methods for Arriving at Clinical Judgments in Peer Evaluation

Anthony A. Cancelli

Fordham University

Running Head: CLINICAL JUDGMENTS

This paper was presented at the convention of the American Educational Research Association in April, 1987. Requests for reprints should be sent to Anthony A. Cancelli, Graduate School of Education, Fordham University, Lincoln Center, New York, NY 10023.

TM 870 344

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Cancelli

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Abstract

The peer evaluation model of university teaching is composed of three domains of behavioral and cognitive skills which, in combination, represent the elements expected of professionally competent teachers. The measurement of competence in each of these domains posed a particularly difficult challenge. Consistent with the underlying philosophy of the model, methods were developed for collecting data in each of the domains. The data base for making judgments is constructed from interviews with the faculty member and his/her students, classroom observations, and a review of the course syllabus and other course materials. Independent clinical judgments based on definitions of expertness are made by each of six reviewers. Following these independent judgments, disagreements are discussed and consensus reached. The purpose of this paper is to describe the methods employed to collect the data and the processes used to arrive at a consensus clinical judgment.

Methods for Arriving at Clinical Judgments in Peer Evaluation

The proposed peer evaluation model of university teaching is composed of three domains of behavioral and cognitive skills which, in combination, represent the elements expected of professionally competent teachers. These domains are: (1) the conception of the professional role for which students are being prepared; (2) instructional design; and (3) the execution of the instructional strategy. The method used to collect data and arrive at judgments concerning performance was developed after the identification of the specific question we wished to ask in each of the domains. The questions identified for each domain are identified below.

Conception of the Professional Role

The idea behind the questions posed in this domain is that a professor is not teaching a course but rather is educating professionals in a program of study. The gist of the questions, then, relate to the professor's understanding of the professionals s/he is educating, the development of the knowledge and skills required for competent functioning, and how s/he contributes to this development by the course under review.

1. Does the instructor have a conception of the professional that s/he is attempting to produce?

2. Does the instructor have an understanding of alternative conceptions that are viable in the profession?

3. Does the instructor have a conception of the knowledge and skills required of the students to become competent professionals?

4. How does the instructor's conception match what the students in the program are currently learning?

5. Are the graduates of the program unique in any way?

6. Does the instructor have a model for the development of the knowledge and skills required?

7. What are the purposes, goals, and/or objectives of the course and why were they chosen? What should the students be able to do at the end of the course that they couldn't do at the beginning?

8. What is the relevance of the goals of the particular course to professional development?

9. How is the course placed among other courses in the program? Is it foundational? Are there skills and knowledge that are prerequisite to those in the course? Are there skills and knowledge that should occur subsequent to those learned in the course?

Instructional Design

The major thrust of the questions asked in this domain is to identify how the professor plans the course to teach the goals identified in the conception domain. It is the purpose of these questions to (1) find out how well thought out the planning is and (2) collect sufficient information to judge the relevance of the type of instruction to the stated purposes.

1. How long has the instructor taught the course? Does the instructor conceive of it as an experimental course? When was the last time the course was revised?

2. Are some goals more important than others? Do all goals apply to all students in the class?

3. What are the instructional activities of the course? How are class meetings conducted? What are the out-of-class activities required?

4. How does the instructor see these activities producing attainment of the goals of the course? For example, if a goal of the course is to develop skills related to solving particular problems, how does the instructor see the instructional strategy s/he has chosen accomplishing this goal?

5. How does the instructor evaluate how well the goals of the course have been achieved?

6. How does the instructor see the methods of evaluation providing the necessary information to judge student attainment of the goals? Does the instructor use feedback to help form the development of the students or is it used strictly for making summative judgments?

7. Is the instructor satisfied with the course? How does s/he judge its success? Are there any changes in the course being considered?

Execution of Instructional Strategy

The purpose of the questions in this domain is to identify how well the professor executes his/her instructional plan.

1. Are the goals of the course clearly stated for the students and do they understand them?

2. Do the students understand the activities and how they relate to the goals of the course?

3. Is the instructional plan followed? If not, why?

4. Are the students as involved in the instructional process as planned?

5. Is the interaction between the instructor and students, open, free, respectful, and supportive on the part of the instructor?

6. Do the students understand what is expected of them and how their performance in the course is to be evaluated?

7. Are the procedures used to evaluate the student consistent with the instructor's stated method?

8. Do the students achieve the goals of the course?

Sources of Data

The data used to answer the questions posed above are collected through:

1. interview with the professor

2. observation of instruction

3. review of the course syllabus, reading list, tests and test scores, student papers, and handouts or other materials used in instruction.

4. interview with the students.

Data collected through interview with the professor is mainly used to answer question in the first two domains. The observation of instruction and interview with the students is designed primarily to answer questions in the third domain. The syllabus and other materials are used to provide information concerning questions in all three domains.

Data Analysis and Judgments of Teaching Competence

Each of the three domains of content are "dimensionalized" to aid the sorting of information by the reviewer. Each of the dimensions within each of the domains is a scale. Together the scales define the salient features of the information to which the reviewer must attend.

Domain I, the "Conception" domain, has three dimensions: (1) Clarity, (2) Model of the Role, and (3) Content. Domain II, the "Instructional Design" domain, has four dimensions: (1) Clarity and Consistency of Objectives, (2) Communication of Objectives, (3) Logic of the Instructional Method, and (4) Relevance and Orderliness of the Evaluation System. Domain III, the "Execution" domain, consists of three dimensions: (1) Structure and Movement, (2) Interaction with Students, and (3) Evaluative Feedback.

The reviewer has a major organizing activity. He or she must take a rich variety of information and sort it first into domains and then into dimensions within the domain. In principle, any unit of data can be sorted into these domains and along these dimensions.

The next task of the reviewer is to sort the data in each of the dimensions into one of three categories. The categories are: (1) below expected performance, (2) standard expected performance, and (3) above expected performance. These categories were chosen to represent the performance expected of a

competent professor. The competent professor is one who has a solid concept of the role for which s/he is preparing students, designs his or her instructional activities to achieve this purpose, and executes the instructional strategies well, including evaluating students with appropriate assessment procedures and relating to students positively and constructively. This represents the midpoint of the dimensions and is labeled standard expected performance. It does not mean average, instead it represents the basically good, solid teaching expected of all professors. Below expected performance represents performance that is less than expected of a good instructor on that dimension.

Conceptually, the units of data available for each dimension are viewed as being normally distributed within the dimension. It is the reviewers job to identify the category in which the midpoint on the distribution of data is located within each dimension. This becomes the score for the professor on that dimension. Each professor, then, receives 10 scores, three in Domain I, four in Domain II and three in Domain III (see Appendix A).

The tasks of the reviewer in analyzing and interpreting the data are aided by a scoring manual. Each of the domains are described, the objective of the domain stated, and critical questions listed. It is the reviewer's answers to these critical questions that comprise the analysis of the data. This provides the bits of information which are then used in making the

judgments. The judgments are aided by a detailed description of what constitutes below, standard, and above expected performance in each of the dimensions. The judgment involves locating the professor at one of the three levels by estimating how closely what has been heard and seen by the reviewer, fits the description of a particular level.

The description is not a checklist. It is a picture. The reviewer constructs a mental image of the professor's performance in each dimension and matches it to this picture. The judgment is essentially holistic. It is not an algebraic sum of pluses and minuses. In making this match, the reviewer must be aware of what s/he is attending to and adhere to the definitions of the categories.

Procedures

The procedures described below are those that are pertinent to the formation of a peer review committee; data collection; and analysis and interpretation. They do not include all of the procedures that should be executed to make the system fair and sensitive to the feelings and rights of the professor being reviewed.

Formation of a Peer Review Committee

The model of peer evaluation presented in this symposium uses the clinical judgments of peers to evaluate the teaching competency of university faculty members. An underlying assumption of the model is that through the use of clinical judgments, valid and reliable decisions may be made if proper steps are taken to address some of the inherent difficulties in their use.

The use of clinical judgments as a method for the peer evaluation system presently described offers a major advantage. The system makes minimal assumptions regarding how instruction should occur. It is left to the professional judgments of the professors to determine how they wish to develop and teach their courses. The system only requires that they be within the bounds of acceptable practice, broadly defined, they do what they say they do, and that there be a cogent rationale for their choices. Thus, the review of each professor is unique and requires decisions based on disparate and often idiosyncratic bits of information. The use of judgments by peers provides a method that is flexible enough to adjust to the unique data base generated in each review.

All decisions require some form of judgment. Decision making is a human activity. No set of data ever makes decisions; people make decisions. As Cronbach (1971) points out, all decisions are prospective. Regardless of the evidence offered

for the valid use of a set of data, one must make judgments regarding their use in the particular setting, for the particular individual, at that particular point in time, and to predict the specific outcomes of interest. Yet, the use of judgments to the extent that the process become a private matter of selecting, analyzing, and weighting data to draw conclusions based, for the most part, on ones own set of private norms requires careful planning. In developing the present system several steps are taken to help insure the integrity of the judgments.

The first involves procedures for selecting who conducts the review. The system requires that several people serve on a committee to conduct the reviews. At a minimum, three reviewers should be used; we used five. Since judgments are so susceptible to biases, faulty understandings and the like, multiple independent judgments need to be arrived at so that inter-judge reliability may be checked.

It is also recommended that the people chosen for the committee be those peers who are recognized for being fair-minded, even-handed, and without extreme points of view. You want the reviewers to have a degree of technical competence and capable of making sound professional judgments after learning the system.

Once the committee is formed, its members must be trained to use the system. The purpose of the training is to calibrate the clinicians; that is, to train the clinicians to function in a

valid and reliable way. It should be clear among the reviewers what data is to be used and the definitions of each of the dimensions and categories. Once the committee arrives at a common understanding of the system, a practice review should be conducted with a faculty volunteer.

Data Collection

Two major events comprise the data collection phase of the process. The first is the faculty interview, the second, the instructional observation.

Faculty interview. As indicated above, the faculty interview is conducted primarily to answer questions in Domains I and II. Before the actual interview, the professor is fully briefed on the process, selects the course that will be used for the review, and provides the course materials to the committee. The committee reviews the materials in advance, and meets prior to the interview to prepare. The committee then meets with the professor during which time they discuss the professors views, goals, and instructional and evaluation strategies according to the Interview Protocol developed for this purpose. The final phase of the interview focuses on what class will be observed.

Instructional observation. Two observers each attend one session of the chosen teaching activity. Before each observation the reviewer meets with the professor to identify what is planned for the class. What are the goals? How do the goals fit into the larger scheme of things? What should the reviewer look for? How will the class be conducted? What is expected of the students? The reviewers job during the observation is to see if the plan is executed as described.

After the class meeting, the reviewer interviews a small sample of the students chosen at random. The interview with the students is designed to collect information concerning the class just observed as well as the course in general. Specifically, the reviewer is interested in finding out the students' understanding of what the professor was trying to communicate and their opinions on the success of the professor's strategy for doing so. Also, the reviewer asks questions regarding how typical the class was, if it is clear to them how the course fits into their studies, and whether or not the professor is respectful and supportive of the students. After the student interview, the reviewer meets with the professor to discuss the professors impressions of the class and to discuss any deviations from the plan that may have occurred.

Data Analysis and Interpretation

After all the data is collected, the committee meets to analyze and interpret the data. Each committee member privately reviews his or her notes and available documents. The chairperson of the committee then reviews the definitions of the dimensions one by one and the members indicates into which category the professor's performance falls. Each reviewer then scores the professor on the scoring sheet (see Appendix A). Using the Scorer Agreement Form (see Appendix B), the chairperson identifies the agreements and disagreements and computes the percentage agreements.

Disagreements are then discussed and a consensus judgment made. The consensus judgment on each dimension is then recorded and supporting statements are written on the form together with recommendations to the professor.

The committee then meets with the professor to provide feedback.

Conclusions

Teaching is an activity that requires frequent and continual modifications. It requires adaptation to an ever changing audience varied in interests, needs, and professional development. The professor's own scholarship and that of others requires frequent modification of content, goals, and the strategies for reaching those goals. In addition to the

variations in courses over time, there is even greater variation from one course a professor teaches to another. This makes the job of evaluating teaching competency elusive to say the least. Consequently, a system designed to evaluate this process must be flexible. It must be sensitive to those variations in teaching that make it such an exciting and challenging activity. Reducing such a vibrant process to a checklist of behaviors performed would reduce teaching to a mechanical act. It may be argued, therefore, that the use of a system that allows peers to use their analytic processes to scan, classify, sort, and resort, the complex and unique sets of data presented in each review is the only way to do justice to teaching.

References

Cronbach, L.J. (1971). Test validity. In R.L. Thorndike (Ed.). Educational measurement (Second Edition) (pp. 443-507). Washington, D.C.: American Council of Education.

Appendix A

SCORING SHEET

Professor Reviewed_____ Date_____ __

Reviewer_____

Check your participation in the review process below:

Present at Interview _____ Observed Teaching _____ # of Obs._____

Read documents _____ Post-Session _____ Student Interview _____

DIRECTIONS: Across from the name of each dimension place a checkmark in the box which best represents your evaluation of the professor's performance on that dimension.

DOMAIN AND DIMENSIONS	BEP	SEP	AEP
DOMAIN I			
CLARITY			
MODEL OF THE ROLE			
CONTENT			
DOMAIN II			
CLARITY AND CONSISTENCY OF OBJECTIVES			
COMMUNICATION OF OBJECTIVES			
LOGIC OF THE INSTRUCTIONAL METHOD			
RELEVANCE AND ORDERLINESS OF EVALUATION SYSTEM			
DOMAIN III			
STRUCTURE AND MOVEMENT			
INTERACTION WITH STUDENTS			
EVALUATIVE FEEDBACK			

SUMMARIZE YOUR CHECKS HERE:

OF BEP'S _____ # OF SEP'S _____ # OF AEP'S _____

YOUR OVERALL EVALUATION: BEP ___ SEP ___ AEP ___ ___
(Check One)

SUPPORTING STATEMENTS:

RECOMMENDATIONS TO THE PROFESSOR:

Appendix B

SCORING AGREEMENT SHEET REVIEW # ____

PROFESSOR _____ DATE __ __

DOMAINS AND DIMENSIONS	REVIEWERS 1	2	3	4	5
DOMAIN I					
Clarity					
Model					
Content					
DOMAIN II					
Clarity and Consistency					
Communication Of Objectives					
Logic of Instr. Method					
Relev.and Ord. of Eval. Sys.					
DOMAIN III					
Structure & Movement					
Interaction					
Eval.Feedbk.					

# of Agreements_____ % Agreement _____